On the wavelet denoising of multifractal functions

A. Fraysse*

Abstract

This paper is devoted to the link between the rate of convergence of functions by thresholding algorithms and the multifractal formalism. Indeed we show that once given its Besov regularity, almost every function in the prevalence sense, is approximated at the minimax rate. We also prove that when the multifractal formalism is fulfilled the corresponding functions are approximated at the minimax rate of convergence.

AMS classification:
62G05, 62G20, 28C20, 46E35

Keywords:
minimax theory, maxiset theory, Besov spaces, prevalence, wavelet bases, multifractal analysis, Frisch-Parisi conjecture

1 Introduction

The increasing interest on nonparametric statistics in the last decades comes from an important need for accurate estimation methods in a wide class of applicative contexts. The aim in this case is to determine estimators of infinite dimensional objects, such as functions for instance. Another important question in this case is on how an estimation procedure can be said to be efficient. The classical approach is thus to determine the minimax rate of

*L2S, SUPELEC, CNRS, University Paris-Sud, 3 rue Joliot-Curie, 91190 Gif-Sur-Yvette, FRANCE. Email: fraysse@lss.supelec.fr.

approximation reached by this estimator. This minimax rate of convergence is obtained by considering the minimal, over all estimation procedures, of the maximal error in a given space. Although minimax theory is widely used in theoretical and practical studies, its definition is subject to controversy. First of all, it allows to merge different estimators by finding their worst performances. Furthermore, in practical studies, the hypothesis of a given space seems too restrictive to be taken into account.

In this paper, we are mainly interested on those two drawbacks and on how they can be relaxed. For this purpose, we consider the classical estimation problem given by the white noise model and we focus on thresholding algorithms, commonly used in practice. This paper is inspired by [9] and [11]. These two papers linked, with different points of view, performances of those algorithms with the frontier of the Besov domain of functions. In [9] it is proved that the generic behaviour of thresholding in a given Besov space coincides with its minimax risk. On the other side, the authors of [11] studied the minimax risk obtained when relaxed the hypothesis of belonging to a Besov ball. They also studied the possible relationship between estimation performances and multifractal analysis. The main purpose of this paper is to merge those two results and to show how minimax performances are related to multifractal behaviour.

Let us in a first time define multifractal analysis and its related tools. This theory was developed to study fully developed turbulent flows. Regularity of those kind of signals changing wildly from point to point it is hardly computable in practice. Hence, rather than studying the values of this regularity, one proposed to study sets of points where it takes a given value H. This leads to the *spectrum of singularities* $d(H)$ which gives for each H the fractal dimension of the set of points where the regularity of a function is exactly H.

The main notion of multifractal analysis is the measure of the regularity which is given by the Hölder exponent.

Definition 1. *Let* $\alpha \geq 0$; *a function* $f : \mathbb{R}^d \to \mathbb{R}$ *is* $C^\alpha(x_0)$ *if for all* $x \in \mathbb{R}^d$ *such that* $\|x - x_0\| \leq 1$ *there exist a polynomial* P *of degree less than* $[\alpha]$ *and a constant* $C > 0$ *such that,*

$$|f(x) - P(x - x_0)| \leq C\|x - x_0\|^\alpha. \tag{1}$$

The Hölder exponent of f at x_0 is

$$h_f(x_0) = \sup\{\alpha : f \in C^\alpha(x_0)\}.$$

As the spectrum of singularities cannot be obtained numerically, in [10] two physicists U. Frisch and G. Parisi proposed the following heuristic formula, the so called multifractal formalism, to obtain this spectrum:

$$d(H) = \inf_{p\in\mathbb{R}}(pH - \xi_f(p) + d). \tag{2}$$

Here $\xi_f(p)$ stands for the L^p modulus of continuity of the function f. This quantity is called the *scaling* function, or scaling exponent of f, and is defined by $\int |f(x+l)-f(x)|^p dx \sim |l|^{\xi_f(p)}$, where $\sim$ means that $\int |f(x+l)-f(x)|^p dx$ is of the order of magnitude of $|l|^{\xi_f(p)}$ when l tends to 0. Numerical estimations and further results about the scaling function and its wavelet decomposition can be found in [1, 2].

It is proved in [15] that for $p \geq 1$, the scaling function $\xi_f(p)$ is closely related to Sobolev or Besov smoothness. It is thus natural to replace the scaling function $\xi_f(p)$ as follows:

$$\text{If } p > 0 \quad \eta_f(p) = \sup\{s : \ f \in B_p^{s/p,\infty}\}. \tag{3}$$

So (2) applied to η_f can at most give the increasing part of the spectrum. Recent results such as [17] retrieve the decreasing part by considering oscillations spaces instead of Besov ones.

Defining, as in [16], an auxiliary function $s(1/p) = \eta_f(p)/p$, the Besov domain of a function f is the set of (q,t) such that $f \in B_{1/q}^{t,1/q}$. The boundary of the Besov domain of f is then given by the graph of $s(q)$. And by Sobolev embeddings, the Besov domain of a function is a convex set. Thus, the functions η satisfying (3) are increasing and concave. Furthermore the auxiliary function s is such that $0 \leq s'(q) \leq d$. These facts lead us to the following definition.

Definition 2. *A function η is admissible if $s(q) = q\eta(1/q)$ is concave and satisfies $0 \leq s'(q) \leq d$. Furthermore it is strongly admissible if $s(0) > 0$.*

Furthermore, see [16], the following proposition entails the properties of admissible functions.

Proposition 1. *Any concave function s satisfying $0 \le s'(q) \le d$ defines the Besov domain of a distribution f.*

Thanks to Proposition 1, to each admissible function η, can be associated a metric space V by taking

$$V = \bigcap_{\varepsilon>0, 0<p<\infty} B_{p,loc}^{(\eta(p)-\varepsilon)/p,p}. \tag{4}$$

One can notice that this topological vector space can be seen as an extension of classical Besov spaces. Indeed, if a function f belongs to a given Besov space $B_{p_0}^{s_0,\infty}$ it also belongs to a space V defined thanks to an admissible function s where:

$$\forall p > 0 \quad s(p) = \begin{cases} s_0 & \text{if } p \le p_0 \\ \frac{d}{p} + s_0 - \frac{d}{p_0} & \text{if } p \ge p_0. \end{cases} \tag{5}$$

The main result of [11] and the main purpose of the present paper is the statistical denoising of functions of V.

In the following, we treat the statistical problem given by the Gaussian white noise model. As in [13], we suppose that we observe Y_t such that

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in (0,1)^d, \tag{6}$$

where dW_t stands for the d-dimensional Wiener measure, n is known and f is the unknown function to be estimated.

The main issue is to determine a good estimation procedure that can retrieve f knowing the observation of Y. Given a pseudo distance $R(.,.)$, the risk function, an estimator $\hat{f}_n$ measurable respectively to Y is minimax on a function space if it attains, at least asymptotically, the minimum value of the maximal risk on this space. Hence given a function space Θ and a closed ball Θ_C of radius $C > 0$, the maximal risk attainable by $\hat{f}_n$ on Θ_C is given by

$$R^n(\hat{f}_n) = \sup_{f \in \Theta_C} \mathbb{E}(R(\hat{f}_n, f)), \tag{7}$$

and the minimax risk is given by

$$R^n(\Theta) = \inf_{\hat{f} \in \mathcal{T}_n} \sup_{f \in \Theta_C} \mathbb{E}(R(\hat{f}_n, f)), \tag{8}$$

where $\mathcal{T}_n$ is the set of estimation procedure measurable respectively to Y.

In [9], it is shown that in a given Besov space the minimax risk for this model is indeed attained by the thresholding estimator on generic sets. Genericity used there is given by the notion of prevalence, that is in an infinite dimensional measure theoretic sense.

The first definition of this "almost every" notion is given by J. Christensen in 1972 see [3, 4, 12].

Definition 3. *Let E be a complete metric vector space. A Borel set $A \subset E$ is Haar-null (or shy) if there exists a compactly supported probability measure μ such that*

$$\forall x \in E, \quad \mu(x + A) = 0. \tag{9}$$

If this property holds, the measure μ is said to be transverse to A.

A subset of E is called Haar-null if it is contained in a Haar-null Borel set. The complement of a Haar-null set is called a prevalent set.

The following results enumerate important properties of prevalence and show that these notions supply a natural generalization of "zero measure" and "almost every" in finite-dimensional spaces, see [3, 4, 12].

Proposition 2.

- *If S is Haar-null, then $\forall x \in E$, $x + S$ is Haar-null.*
- *If* $\dim(E) < \infty$, *S is Haar-null if and only if* $\text{meas}(S) = 0$ *(where* meas *denotes the Lebesgue measure).*
- *Prevalent sets are dense.*
- *The intersection of a countable collection of prevalent sets is prevalent.*
- *If* $\dim(E) = \infty$, *compact subsets of E are Haar-null.*

The fact that a set is prevalent or not is independant of the chosen transverse measure. Several kinds of transverse measures, such as the one induces by the sample paths of a random process or the Lebesgue measure on a finite dimensional subspace can thus be used in this context, involving the same results. However, an important drawback of this theory, which is shared by a lot of genericity results, is that it does not allow to characterize element of generic sets.

The remaining of this paper is devoted to the thresholding estimation in the topological vector space V. In order to define this estimation procedure one has to define a basis of the involved space. In our context, this basis is given by wavelet bases.

1.1 Wavelet expansion

The wavelet thresholding was introduced in [7] as a denoising tool in transform domains. The starting point is that in a wavelet decomposition of noisy data, small coefficients should correspond to noise. It is thus natural to remove them in order to recover the signal. Wavelets take here an important role, as they are well localized in time and in frequency domains. Furthermore, it also provides a good tool for regularity criterion.

For r large enough there exists 2^d-1 functions $\psi^{(i)}$ with compact support and which belong to C^r, see [6]. Each $\psi^{(i)}$ has r vanishing moments and the set of functions $\{\psi^{(i)}_{j,k} = 2^{dj/2}\psi^{(i)}(2^j. - k), \quad j \in \mathbb{Z},\ k \in \{0, \dots, 2^j-1\}^d, i \in \{1, ..., 2^d-1\}\}$, which are called wavelets, forms an orthonormal basis of $L^2([0,1]^d)$. It is also noticed in [20] that wavelets provide unconditional bases of $L^p([0,1]^d)$ as far as $1 < p < \infty$.

Thus any function $f \in L^p([0,1]^d)$ can be written as

$$f(x) = \sum_{i,j,k} c^{(i)}_{j,k}\psi^{(i)}_{j,k}(x) \tag{10}$$

where

$$c^{(i)}_{j,k} = 2^{jd/2}\int f(x)\psi^{(i)}(2^jx-k)dx. \tag{11}$$

As the collection of $\{2^{dj/2}\psi^{(i)}(2^j. - k), \quad j \in \mathbb{N},\ k \in \{0, \dots, 2^j-1\}^d, i = 1, ..., 2^d-1\}$ form an orthonormal basis of $L^2([0,1]^d)$, observing the whole trajectory of Y_t in (6) is equivalent to observe its wavelet coefficients $(y_{j,k})_{j\in\mathbb{N},k\in\{0,\dots,2^j-1\}^d} \in \ell^2(\mathbb{N}^d)$ such that $\forall j,k$,

$$y_{j,k} = \theta_{j,k} + \frac{1}{\sqrt{n}}v_{j,k}, \tag{12}$$

where $y_{j,k} = \int \psi^{(i)}_{j,k}dY$, $v_{j,k}$ are i.i.d. Gaussian random variables and $(\theta_{j,k})$ is the sequence to be estimated. One can notice that we stand in an isotropic problem. Thus the direction of the wavelets is not involved and we omit in the following the directional index i.

Furthermore wavelets are useful as they provide a simple characterization of Besov spaces.

Homogeneous Besov spaces are characterized, for $p, q > 0$ and $s \in \mathbb{R}$, by

$$f \in B_p^{s,q}([0,1]^d) \iff \exists C > 0 \ \sum_j \left(\sum_{j \geq 0, k \in \{0, \ldots 2^j - 1\}^d} |c_{j,k}|^p 2^{(sp - d + \frac{pd}{2})j} \right)^{q/p} \leq C. \tag{13}$$

This characterization is independent from the chosen wavelet as soon as it is C^r and has r vanishing moments with $r \geq s$.

Another important function spaces, studied in the following is given by the weak Besov spaces defined in [5]. These spaces are are subsets of Lorentz spaces, and constitute a larger class than Besov spaces.

Definition 4. *Let $0 < r < p < \infty$. We say that a function $f = \sum_{j,k} c_{j,k} \psi_{j,k}$ belongs to $W(r,p)$ if and only if*

$$\sup_{\lambda > 0} \lambda^r \sum 2^{j(\frac{dp}{2} - d)} \sum_k \mathbb{1}_{\{|c_{j,k}| > \lambda\}} < \infty. \tag{14}$$

A fast calculation shows that the space $W(r,p)$ contains homogeneous Besov spaces $B_r^{\beta,\infty}$ for $\beta \geq \frac{d}{2}(\frac{p}{r} - 1)$.

From the localization property of wavelet bases, it also provides a powerful tool in multifractal analysis, thanks in particular to the following proposition from [15].

Proposition 3. *Let x be in $\mathbb{R}^d$. If f is in $C^\alpha(x)$ then there exists $c > 0$ such that for each (j,k):*

$$|c_{j,k}| \leq c 2^{-(\alpha + \frac{d}{2})j} (1 + |2^j x - k|)^\alpha. \tag{15}$$

Unfortunately this condition is not a characterization. If for any $\varepsilon > 0$, a function does not belongs to $C^\varepsilon(\mathbb{R}^d)$ one cannot express its pointwise Hölder regularity in term of condition on wavelet coefficients. However, the following proposition from [14] allows to deduce a lower bound of the Hölder exponent thanks to wavelets properties.

Proposition 4. *Let $x_0 \in \mathbb{R}^d$ be fixed and $s > 0$. Let $f = \sum_{j,k} c_{j,k}\psi_{j,k}$ be such that*

$$\exists s' > 0 \qquad \sup_{j\in\mathbb{Z},k\in\mathbb{Z}^d} \quad |c_{j,k}| \le 2^{-(s+\frac{d}{2})j}(1+|k-2^j x_0|)^{s'} < \infty. \tag{16}$$

Then $f \in C^s(x_0)$.

There is thus a logarithmic loss between (15) and the Hölder exponent. Note that recent results showed that an equivalent condition on wavelet leaders provides a caracterization of the regularity [17].

Let us now define the statistical estimation that we deal with. This procedure, classical in terms of signal estimation is given by the "hard" wavelet thresholding such as given in [7].

Definition 5. *The wavelet hard thresholding estimator is defined by*

$$\hat{f}_n^T(x) = \sum_{j=0}^{j(n)} \sum_k \beta_{j,k}^T \psi_{j,k}(x). \tag{17}$$

Here the weights are given by:

$$\beta_{j,k}^T = y_{j,k} \mathbb{1}_{\{|y_{j,k}| \ge \kappa t_n\}}. \tag{18}$$

where

$$t_n = \sqrt{\frac{\log n}{n}}, \tag{19}$$

is the universal threshold, $j(n)$ is such that

$$2^{-j(n)} \le \frac{\log n}{n} < 2^{-j(n)+1}$$

and κ is a constant large enough.

2 Statement of main results

The purpose of this paper is twofold. First we establish which rate of approximation by wavelet thresholding is generic in a space defined as in (4).

In a second time, we link this result with the multifractal formalism.

In the rest of this paper, we fix η to be an admissible function and V the metric space defined by:

$$V = \bigcap_{\varepsilon>0, 0<p<\infty} B^{(\eta(p)-\varepsilon)/p,p}_{p,loc}, \tag{20}$$

This set V can also be written as a countable intersection of $B^{(\eta(p_n)-\varepsilon_n)/p_n,p_n}_{p_n,loc}$

Note that V is a topological vector space. For $p<1$ Besov spaces are only quasi-Banach spaces, as the triangle inequality is only satisfied up to a constant and V is not a Banach space but a complete metric space. Indeed, if $p \geq 1$ we take for distance between two functions f and g in $B^{s,q}_p$:

$$d(f,g) = \sum_{j\geq 0} \left(\sum_{k\in\{0,\dots,2^j-1\}^d} \left|(c_{j,k}-d_{j,k})2^{(s-\frac{d}{p}+\frac{d}{2})j}\right|^p \right)^{\frac{q}{p}}$$

where $c_{j,k}$ are the wavelet coefficients of f and $d_{j,k}$ are those of g.
If $p<1$ we consider:

$$d(f,g) = \left(\sum_{j\geq 0} \left(\sum_{k\in\{0,\dots,2^j-1\}^d} |(c_{j,k}-d_{j,k})2^{(s-\frac{d}{p}+\frac{d}{2})j}|^p \right)^{\frac{q}{p}} \right)^{\frac{\min(p,q)}{q}}.$$

Finally, we obtain a distance in V by taking:

$$\forall f,g \in V \quad d(f,g) = \sum_n 2^{-n} \frac{d_n(f,g)}{1+d_n(f,g)}$$

where d_n denotes the distance in $B^{(\eta(p_n)-\varepsilon_n)/p_n,p_n}_{p_n,loc}$. With this distance V is clearly a complete space.

Furthermore, as in [11], for each $r \geq 1$ we define $p^*(r)$ as the solution, if it exists, of

$$s(1/p) = \frac{d}{2}\left(\frac{r}{p}-1\right). \tag{21}$$

Note that a sufficient condition for existence of $p^*(r)$ is that $s'(\infty) < \frac{dr}{2}$. Let us now state our main result. In the following $a_n \approx b_n$ means that $\frac{\log(a_n)}{\log(b_n)} \to 1$.

Theorem 1. *Let* $1 \le r \le \infty$ *such that* $s'(\infty) < \frac{dr}{2}$, *then in the context of* (6), *for almost every* $f \in V$ *and for the thresholding algorithm* $\hat{f}_n^T$

$$\mathbb{E}\|\hat{f}_n^T - f\|_{L^r}^r \approx \left(\frac{n}{\log n}\right)^{-\alpha r} \tag{22}$$

where

$$\alpha = \frac{s(1/p^*(r))}{2s(1/p^*(r)) + d}. \tag{23}$$

As one can see, this result is similar to the minimax result of [11].

Thanks to Theorem 1, we can also prove the following result.

Theorem 2. *For every function* $f \in V$ *satisfying*

$$\forall H \in \left[s(0), \frac{d}{p_c}\right] \quad d_f(H) = \inf_{p \ge p_c} (pH - \eta_f(p) + d) \tag{24}$$

where p_c *is the only critical point such that* $\eta(p_c) = d$, *the rate of estimation of* f *by wavelet thresholding in the context of* (6) *satisfies:*

$$\mathbb{E}\|\hat{f}_n^T - f\|_{L^r}^r \approx \left(\frac{n}{\log n}\right)^{-\alpha r}, \tag{25}$$

where

$$\alpha = \min_{H \in \mathrm{supp}(d)} \frac{H + (d - d(H))/r}{2H + d}. \tag{26}$$

Stated differently, in this second theorem we see how the multifractal behaviour is linked to estimation performances. Actually, one can check in [8] and in Theorem 1 that both multifractal and minimax properties represent characteristic behaviour in Besov spaces. The question raised thus is to understand which one is the most "generic". The answer given here is that the multifractal formalism actually entails minimax rate of estimation by wavelet thresholding.

3 Proof of Theorems 1 and 2

In order to prove the main results of this paper we have to introduce an useful tool given by the maxiset theory introduced in [5, 18, 19] as an alternative way to compare different estimation procedures.

3.1 Maxiset theory

The main idea of this theory is to consider the maximal space on which an estimator reach a given rate, instead of searching an optimal rate for a given space.

Definition 6. *Let ρ be a risk function and $(v_n)_{n\in\mathbb{N}}$ a sequence such that $v_n \to 0$. For $\hat{f}_n$ an estimator measurable with respect to Y given by* (6)*, the maximal space associated to ρ, v_n and a constant T is given by*

$$MS(\hat{f}_n, \rho, v_n, T) = \left\{ f;\ \sup_n v_n^{-1}\mathbb{E}(\rho(\hat{f}_n, f)) < T \right\}. \tag{27}$$

The maxiset associated with the thresholding estimation procedure is given by a weak Besov space as proved in [5].

Proposition 5. *Let $1 \leq p < \infty$, $1 \leq r < \infty$, $s > \frac{d}{r}$ and $\tilde{\alpha} \in (0,1)$. Let $\hat{f}_n^T$ be the estimator defined by* (17) *and* (18)*. Then for every f we have the following equivalence:*
$\exists K > 0$ such that $\forall n > 0$,

$$\mathbb{E}\|\hat{f}_n^T - f\|_p^p \leq K\left(\sqrt{n\log(n)^{-1}}\right)^{-\tilde{\alpha}p} \tag{28}$$

if and only if $f \in B_p^{\tilde{\alpha}/2,\infty} \cap W((1-\tilde{\alpha})p, p)$.

Furthermore the following proposition from [8] gives also a key argument to prove Theorem 1.

Proposition 6. *For almost every f in V, we have:*

$$\forall 0 < p < \infty \quad \eta_f(p) = \eta(p). \tag{29}$$

3.2 Generic rate for thresholding algorithms

Let us now prove Theorem 1. For this purpose we use the following proposition from [11], which gives the upper bound.

Proposition 7.

$$\mathbb{E}(\|\hat{f}_n^T - f\|_r^r) \leq c\left(\frac{n}{\log n}\right)^{-\frac{s(1/p^*(r))}{2s(1/p^*(r))+d}r}. \tag{30}$$

Let us now determine the lower bound. For this purpose, let $r \geq 0$ be fixed and $\hat{f}_n^T$ the estimator given in Definition 5.

Let us turn out our attention to the minimax rate of convergence. For this purpose, we write in the following

$$\tilde{\alpha}(r) = \frac{2s(1/p^*(r))}{2s(1/p^*(r)) + d}. \tag{31}$$

For every values of $\tilde{\alpha}(r)$, let $0 < \varepsilon < 1 - \tilde{\alpha}(r)$ be fixed, and $M(\varepsilon)$ be the set defined by

$$M(\varepsilon) = \left\{ f \in V;\ \exists c > 0\ \forall n \in \mathbb{N},\ \mathbb{E}(\|\hat{f}_n^T - f\|_{L^r}^r) < c\sqrt{\frac{n}{\log n}}^{-(\tilde{\alpha}(r)+\varepsilon)r} \right\}. \tag{32}$$

Thanks to Proposition 5, this set $M(\varepsilon)$ is embedded in $B_p^{\frac{\tilde{\alpha}(r)+\varepsilon}{2},\infty} \cap W((1-\tilde{\alpha}(r) - \varepsilon)r, r)$.

Let us now prove that $W((1-\tilde{\alpha}(r)-\varepsilon)r, r)$ is a Haar null Borel set in V.

For this purpose, we consider as a transverse measure, the Lebesgue measure on the set generated by the function g defined thanks to its wavelet coefficients:

$$d_{j,k} = \frac{2^{a(j,k)}}{j^a} \tag{33}$$

where $a = a_j = \log j$ and

$$a(j,k) = \inf_p \left(\frac{d(j-J) - \eta(p)j - pdj/2}{p} \right). \tag{34}$$

In this definition, $0 \leq J \leq j$ and $K \in \{0, \dots, 2^J - 1\}^d$ are such that

$$\frac{K}{2^J} = \frac{k}{2^j} \tag{35}$$

is an irreducible fraction. As it can be seen in [16], this function g belongs to V. Let $f \in V$ be an arbitrary function and consider the affine subset

$$M = \{\alpha \in \mathbb{R}\ \ f + \alpha g \in W((1-\tilde{\alpha}(r) - \varepsilon)r, r)\}. \tag{36}$$

Suppose that there exist two points α_1 and α_2 in M. Thus $f+\alpha_1 g-(f+\alpha_2 g)$ belongs to $W((1-\tilde{\alpha}-\varepsilon)r,r)$, and therefore there exists $c>0$ such that

$$\|f+\alpha_1 g-(f+\alpha_2 g)\|_{W((1-\tilde{\alpha}-\varepsilon)r,r)} = \|(\alpha_1-\alpha_2)g\|_{W((1-\tilde{\alpha}-\varepsilon)r,r)} \leq c. \quad (37)$$

A fast calculation shows that for each $0\leq t\leq r$,

$$\forall \alpha>0,\quad \|\alpha g\|_{W(t,r)} = \alpha^t \|g\|_{W(t,r)} \quad (38)$$

and we just have now to determine $\|g\|_{W((1-\tilde{\alpha}-\varepsilon)r,r)}$. Thanks to equation (14), this is equivalent to determine for every $t>0$, the value of

$$\|g\|_{W((1-\tilde{\alpha}-\varepsilon)r,r)} = 2^{-(1-\tilde{\alpha}-\varepsilon)rt}\sum_{j\geq 0} 2^{j(\frac{dr}{2}-d)}\sum_k \mathbb{1}_{\{d_{j,k}>2^{-t}\}}$$

But by definition of g, when $d_{j,k}>2^{-t}$ we have, by setting $q=1/p$ in (34)

$$\frac{2^{a(j,k)}}{j^{a_j}} > 2^{-t} \Leftrightarrow \sup_q (qd(J-j)+s(q)j+jd/2) \leq t,$$

which implies that

$$\forall q>0 \quad J\leq \frac{1}{qd}(t-s(q)j-\frac{dj}{2})+j,$$

in particular when $q=\frac{1}{p^*}$ where p^* satisfies (21),

$$J\leq \frac{p^*}{d}(t-s(1/p^*)j-\frac{dj}{2})+j, \quad (39)$$

Note that J must be positive thus

$$j(s(1/p^*)-\frac{d}{p^*}+\frac{d}{2})\leq t. \quad (40)$$

We denote by $\tilde{t}=\frac{t}{s(1/p^*)-\frac{d}{p^*}+\frac{d}{2}}$ and by $\tilde{\tilde{t}}=\frac{t}{s(1/p^*)+\frac{d}{2}}$. From definition of weak Besov spaces, we have for every $t>0$,

$$\begin{aligned}
\|g\|_{W((1-\tilde{\alpha}-\varepsilon)r,r)} &\geq 2^{-(1-\tilde{\alpha}-\varepsilon)rt}\sup_{0\leq j\leq \tilde{t}} 2^{j(\frac{dr}{2}-d)}\sum_{J=0}^{j\wedge[\frac{p^*}{d}(t-s(1/p^*)j-\frac{dj}{2})+j]} 2^{dJ}\\
&\geq 2^{-(1-\tilde{\alpha}-\varepsilon)rt}\sup\left(\sup_{0\leq j\leq\tilde{\tilde{t}}} 2^{j(\frac{dr}{2}-d)}\sum_{J=0}^{j}2^{dJ},\ \sup_{\tilde{\tilde{t}}\leq j\leq\tilde{t}} 2^{j(\frac{dr}{2}-d)}\sum_{J=0}^{[\frac{r}{d}(t-js(1/p^*)-\frac{jd}{2})+j]} 2^{dJ}\right)\\
&\geq \frac{2^{-(1-\tilde{\alpha}-\varepsilon)rt}}{2^d-1}\sup\left(\sup_{0\leq j\leq\tilde{\tilde{t}}} 2^{\frac{drj}{2}}(1-2^{-jd}),\ \sup_{\tilde{\tilde{t}}<j\leq\tilde{t}} 2^{j(\frac{dr}{2}-d)}(2^{p^*t}2^{-jp^*(s(1/p^*)+\frac{d}{2}-\frac{d}{p^*})}-1)\right)
\end{aligned}$$

Merging this result with (37) together with (38), we obtain that, if there exist α_1 and α_2 in M then they satisfy that for every $t \geq 0$ and $0 \leq j \leq \tilde{t}$,

$$|\alpha_1-\alpha_2|^{(1-\tilde{\alpha}-\varepsilon)r} \leq \inf\left(\frac{c2^{(1-\tilde{\alpha}-\varepsilon)rt}}{\sup_{0\leq j\leq \tilde{\tilde{t}}} 2^{\frac{drj}{2}}|1-2^{-jd}|}, \frac{c2^{(1-\tilde{\alpha}-\varepsilon)rt}}{\sup_{\tilde{\tilde{t}}<j\leq \tilde{t}} 2^{j(\frac{dr}{2}-d)}|2^{p^*t}2^{-jp^*(s(1/p^*)+\frac{d}{2}-\frac{d}{p^*})}-1|}\right) \tag{41}$$

But p^* is such that $r(s(1/p^*)+\frac{d}{2}-\frac{d}{p^*}) = \frac{dr}{2}-d$. Thus, for t large enough

$$\sup_{\tilde{\tilde{t}}<j\leq \tilde{t}} 2^{j(\frac{dr}{2}-d)}|2^{p^*t}2^{-jp^*(s(1/p^*)+\frac{d}{2}-\frac{d}{p^*})}-1| \sim 2^{rt}.$$

And,

$$\tilde{\alpha} = \frac{2s(1/p^*)}{2s(1/p^*)+d} = 1-\frac{p^*}{r}.$$

Therefore,

$$|\alpha_1-\alpha_2|^{(1-\tilde{\alpha}-\varepsilon)r} \leq c2^{-\varepsilon r} \tag{42}$$

As $1-\tilde{\alpha}-\varepsilon > 0$, it can be deduced from equations (42) that for t large enough, M is of vanishing Lebesgue measure and $W((1-\tilde{\alpha}-\varepsilon)r, r)$ is an Haar null set in $B_p^{s,\infty}$.

Thanks to invariance under inclusion, we have obtained that for every $\varepsilon > 0$, the set of functions f in V such that

$$\exists c>0\ \forall n\in\mathbb{N},\ \mathbb{E}(\|\hat{f}_n^T - f\|_{L^r}^r) < c\sqrt{\frac{n}{\log n}}^{-(\alpha(s)+\varepsilon)p} \tag{43}$$

is a Haar null set.

Taking the countable union of those sets over a decreasing sequence $\varepsilon_n \to 0$, and considering the complementary we obtain that for almost every function in V,

$$\liminf_{n\to\infty} \frac{\log(\mathbb{E}(\|\hat{f}_n^L - f\|_{L^r}^r))}{-r\log n} \leq \alpha(s). \tag{44}$$

Which induces the expected result.

3.3 Multifractal results in a weak Besov space

As it can be seen in the previous proof, Theorem 1 is based on maxiset theory and estimation behaviour in weak Besov spaces. In the following we determine the multifractal behaviour of function in a weak Besov space.

Proposition 8. *Let $0 \le t \le r$ be fixed and $f \in W(t,r)$. For $D > 0$ and $\beta = \frac{D-d}{t} - \frac{d}{2} + \frac{dr}{2t}$, we have*

$$\mathcal{H}_D\{x : \quad f \notin C^\beta(x)\} = 0,$$

where $\mathcal{H}_D$ denotes the D-dimensional Hausdorff measure.

Corollary. *Let $0 \le t \le r$ be fixed. Then for every $f \in W(t,r)$, and for all $H \in \left[\frac{dr}{2t} - \frac{d}{t} - \frac{d}{2}, \frac{dr}{2t} - \frac{d}{2}\right]$,*

$$d(H) \le tH - \frac{dr}{2t} + \frac{dt}{2} + d. \tag{45}$$

Proof. The first step of the proof consists in the construction of a set $E \subset \mathbb{R}^d$ such that $\mathcal{H}_D(E)$ vanishes.

Let $f = \sum c_{j,k}\psi_{j,k}$ be in $W(t,r)$ thus :

$$\exists c > 0 \text{ such that } \forall i > 0 \ \forall j \quad \sum_{k\in\{0,\dots,2^j-1\}^d} 2^{-ti}2^{j(\frac{dr}{2}-d)}\mathbb{1}_{\{|c_{j,k}|>2^{-i}\}} \le c\varepsilon_j,$$

where $(\varepsilon_j) \in l^1(\mathbb{N})$. This implies that

$$\exists c > 0 \text{ such that } \forall i > 0 \ \forall j \quad \sum_{k\in\{0,\dots,2^j-1\}^d} 2^{j(\frac{dr}{2}-d)}|c_{j,k}|^t\mathbb{1}_{\{|c_{j,k}|>2^{-i}\}} \le c\varepsilon_j. \tag{46}$$

Let us denote $d_{j,k} = (|c_{j,k}|2^{\frac{d}{t}j(\frac{r}{2}-1)})^{\frac{t}{D}}$. Let $B_{j,k}$ be the ball centered at $\frac{k}{2^j}$ and of size $d_{j,k}$. Thus (46) entails

$$\forall j \quad \sum_{k\in\{0,\dots,2^j-1\}^d} |\operatorname{diam}(B_{j,k})|^D = \sum_{k\in\{0,\dots,2^j-1\}^d} |d_{j,k}|^D$$

$$\forall j \quad \sum_{k\in\{0,\dots,2^j-1\}^d} (|c_{j,k}|2^{\frac{d}{t}(\frac{r}{2}-1)j})^t \le c\varepsilon_j$$

As a consequence, we obtain that $\sum_k \operatorname{diam}(B_{j,k})^D \to_{j\to\infty} 0$. Let us now denote by

$$E = \limsup_j \bigcup_k B_{j,k}.$$

We can deduce from (46) that

$$\mathcal{H}_D(E) = 0.$$

Let us assume now that $x \notin E$. There exists l such that $\forall j \geq l$, $\forall k$ $x \notin B_{j,k}$ so that:

$$|x - \frac{k}{2^j}| \geq |d_{j,k}|$$

hence

$$|c_{j,k}| \leq c 2^{\frac{d}{t}j(\frac{r}{2}-1)-\frac{D}{t}j} |2^j x - k|^{\frac{D}{t}}. \tag{47}$$

By setting $\beta = \frac{D-d}{t} - \frac{d}{2} + \frac{dr}{2t}$, we deduce from Proposition 4 that $f \in C^\beta(x)$.

The corollary is straightforward, by noticing that $0 \leq D \leq d$.

□

Let us now turn out our attention to the proof of Theorem 2. By setting $t = (1-\alpha)r$, $0 \leq \alpha \leq 1$ in the previous proposition we obtain that for each $f \in W((1-\alpha)r, r)$:

$$\forall H \in \left[\frac{d\alpha}{2(1-\alpha)} - \frac{1}{(1-\alpha)r}, \frac{d\alpha}{2(1-\alpha)}\right] \quad d(H) \leq (1-\alpha)rH - \frac{dr\alpha}{2} + d.$$

Furthermore, from [16] we have for every $f \in V$,

$$\forall H \in \left[s(0), \frac{d}{p_c}\right] \quad d(H) \leq \inf_{p \geq p_c} (pH - \eta(p) + d). \tag{48}$$

And in [11], the following identity is given:

$$\frac{s(1/p^*)r}{2s(1/p^*) + d} = \inf_{s(0) \leq H \leq \frac{d}{p_c}} \frac{rH + d - d(H)}{2H + d}, \tag{49}$$

as soon as (24) is satisfied.

Let $\varepsilon > 0$ be fixed and $M(\varepsilon)$ be the set defined by (32). As $M(\varepsilon) \subset W((1-\alpha-\varepsilon)r, r)$, for every $f \in M(\varepsilon)$ we have for every $H \in \left[\frac{d(\alpha-\varepsilon)}{2(1-\alpha+\varepsilon)} - \frac{1}{(1-\alpha+\varepsilon)r}, \frac{d(\alpha-\varepsilon)}{2(1-\alpha+\varepsilon)}\right]$

$$d(H) \leq (1-\alpha-\varepsilon)rH - \frac{dr(\alpha-\varepsilon)}{2} + d.$$

By taking $\alpha = \frac{2s(1/p^*)}{2s(1/p^*)+d}$, we obtain

$$\forall H \in \left[s(\frac{1}{p^*}) - \frac{d}{p^*}, s(\frac{1}{p^*})\right], \; d(H) \leq p^*H - p^*s(1/p^*) + d.$$

Thus,

$$\forall \varepsilon > 0, \quad M(\varepsilon) \subset \{f \in V,\; d(H) < p^*H - p^*s(1/p^*) + d\},$$

and taking the complementary,

$$\{f \in V,\; d(H) = \inf_{p \geq p_c}(pH - ps(1/p) + d\} \subset \{f \in V,\; \liminf_{n\to\infty} \frac{\log(\mathbb{E}(\|\hat{f}_n^L - f\|_{L^r}^r))}{-r\log n} \leq \alpha(s)\}.$$

Furthermore, in [11] we have that as soon as the multifractal formalism is fulfilled,

$$\alpha(s) = \inf_{H \in \mathrm{supp}(d)} \frac{H - (d - d(H))/r}{2H + d}.$$

Which entails Theorem 2.

References

[1] P. Abry, *Ondelettes et turbulences. Multirésolutions, algorithmes de décomposition, invariance d'échelle et signaux de pression*, Nouveaux Essais. Paris: Diderot , 1997.

[2] A. Arneodo, E. Bacry, and J.F. Muzy, *The thermodynamics of fractals revisited with wavelets*, Physica A. **213** (1995), 232–275.

[3] Y. Benyamini and J. Lindenstrauss, *Geometric nonlinear functional analysis. Volume 1*, Colloquium Publications. American Mathematical Society (AMS), 2000.

[4] J.P.R. Christensen, *On sets of Haar measure zero in Abelian Polish groups*, Israel J. Math. **13** (1972), 255–260.

[5] A. Cohen, R. DeVore, G. Kerkyacharian, and D. Picard, *Maximal spaces with given rate of convergence for thresholding algorithms*, Appl. Comput. Harmon. Anal. **11** (2001), no. 2, 167–191.

[6] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure and App. Math. **41** (1988), 909–996.

[7] D. Donoho, *Asymptotic minimax risk for sup-norm loss: solution via optimal recovery*, Probab. Theory Related Fields **99** (1994), no. 2, 145–170.

[8] A. Fraysse, *Generic validity of the multifractal formalism*, SIAM J. Math. Anal. **37** (2007), no. 2, 593–607.

[9] ———, *Why minimax is not that pessimist*, ESAIM PS (2012).

[10] U. Frisch and G. Parisi, *On the singularity structure of fully developed turbulence*, Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics, 1985, pp. 84–88.

[11] A. Gloter and M. Hoffmann, *Nonparametric reconstruction of a multifractal function from noisy data*, Probab. Theory Relat. Fields (2009).

[12] B. Hunt, T. Sauer, and J. Yorke, *Prevalence: A translation invariant "almost every" on infinite dimensional spaces*, Bull. A.M.S **27** (1992), no. 2, 217–238.

[13] I. A. Ibragimov and R. Z. Hasminski, *Statistical estimation*, Applications of Mathematics, vol. 16, Springer-Verlag, 1981.

[14] S. Jaffard, *Pointwise smoothness, two-microlocalisation and wavelet coefficients*, Pub. Mat. **35** (1991), 155–168.

[15] ———, *Multifractal formalism for functions*, SIAM J. Math. Anal **28** (1997), 944–970.

[16] ———, *On the Frisch-Parisi conjecture*, J. Math. Pures Appl **79** (2000), 525–552.

[17] S. Jaffard, B. Lashermes, and P. Abry, *Wavelet leaders in multifractal analysis*, Wavelet Analysis and Applications (Tao Qian, Mang I Vai, and Yuesheng Xu, eds.), Applied and Numerical Harmonic Analysis, 2007, pp. 201–246.

[18] G. Kerkyacharian and D. Picard, *Thresholding algorithms, maxisets and well-concentrated bases*, Test **9** (2000), no. 2, 283–344, With comments, and a rejoinder by the authors.

[19] ———, *Minimax or maxisets?*, Bernoulli **8** (2002), no. 2, 219–253.

[20] Y. Meyer, *Ondelettes et opérateurs*, Hermann, 1990.